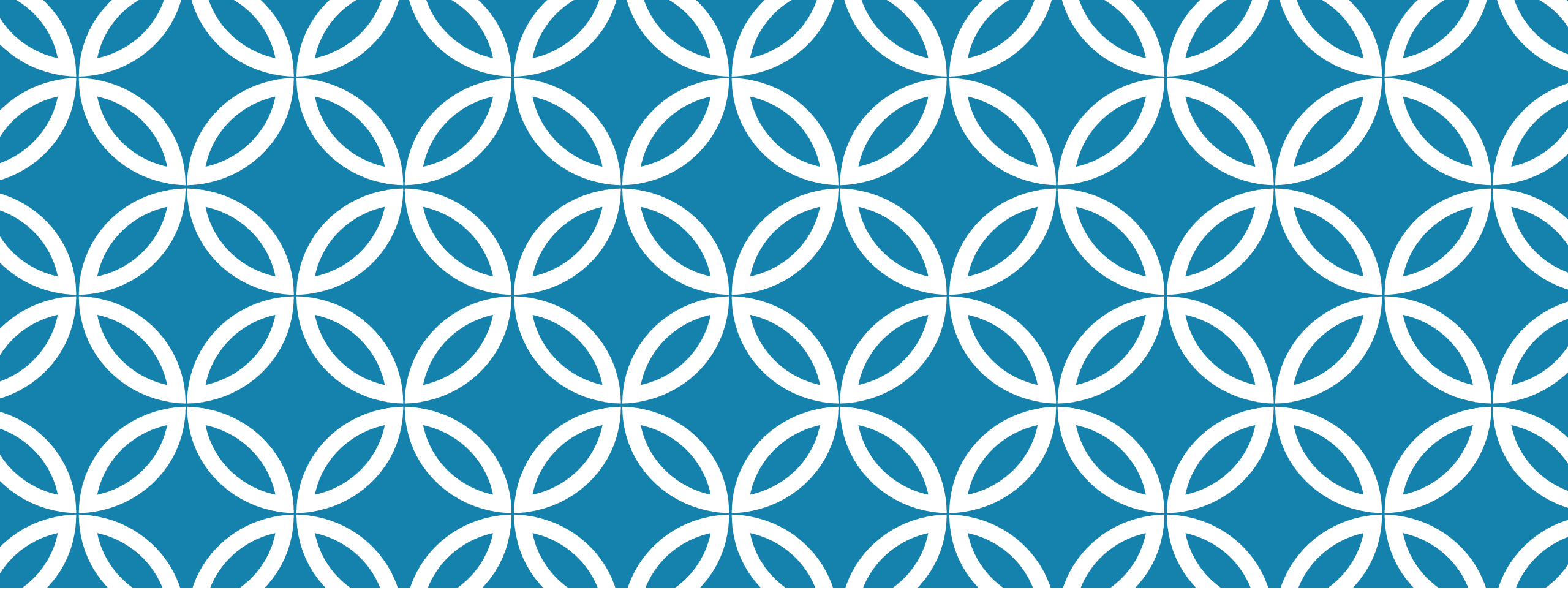


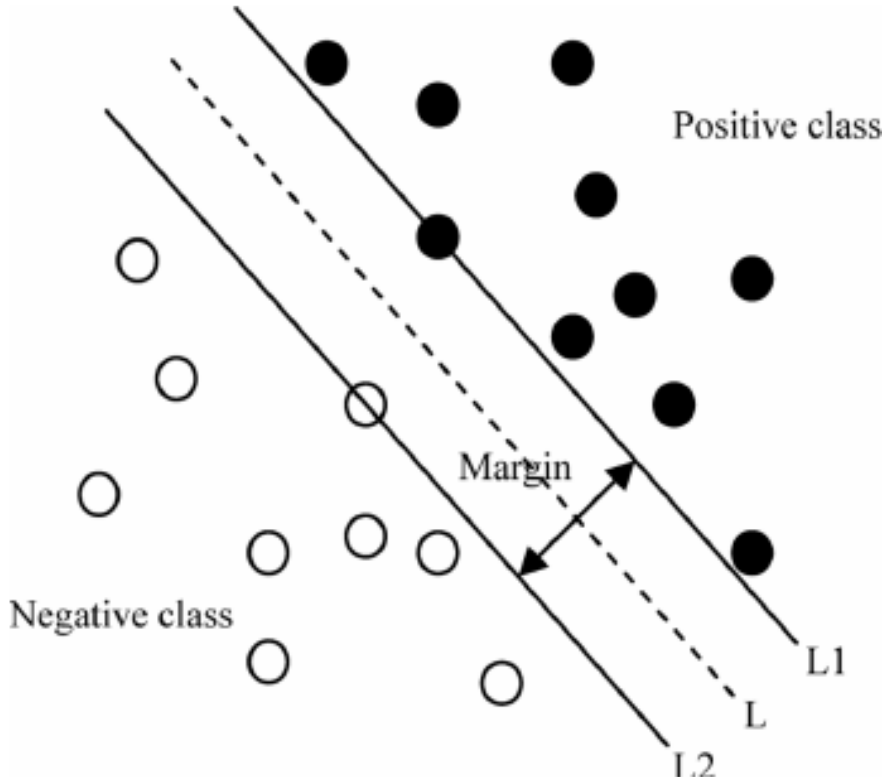
FOREGROUND-BACKGROUND CLASSIFICATION IN HIGH-RESOLUTION BRAIN SCANS

Kartik Chanana
Anirudh Devgun
Sameer Motwani

MODELS TESTED

1. **Support Vector Machine**

SVM is a supervised machine learning algorithm for classification where the dataset teaches SVM about the classes so that SVM can classify any new data. It works by classifying the data into different classes by finding a line (hyperplane) which separates the training data set into classes. As there are many such linear hyperplanes, SVM algorithm tries to maximize the distance between the various classes that are involved and this is referred as margin maximization.



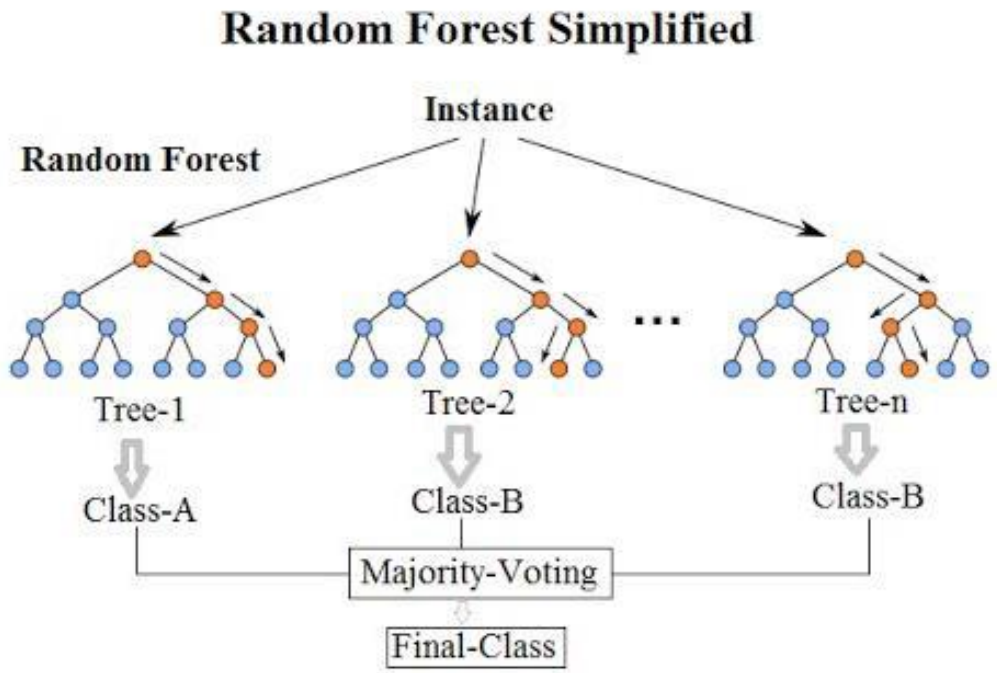
ACCURACY NUMBERS FOR DIFFERENT PARAMETERS

Number of Iterations	Reg Factor	Accuracy	Area under ROC curve
1	1	99.1479	0.5
50	1	99.1479	0.5
75	1	99.1479	0.5
100	1	99.1480	0.500179
50	0.1	99.1479	0.5

MODELS TESTED

2. Random Forest

Random Forest is the go to machine learning algorithm that uses a bagging approach to create a bunch of decision trees with random subset of the data. A model is trained several times on random sample of the dataset to achieve good prediction performance from the random forest algorithm. In this ensemble learning method, the output of all the decision trees in the random forest, is combined to make the final prediction. The final prediction of the random forest algorithm is derived by polling the results of each decision tree or just by going with a prediction that appears the most times in the decision trees.



ACCURACY NUMBERS FOR DIFFERENT PARAMETERS

Number of Trees	Max Depth	Number of Bins	Accuracy	Area under ROC curve
10	4	32	99.5797	0.779993
5	4	32	99.5214	0.739926
10	10	32	99.7375	0.894846
5	10	32	99.6892	0.862126
20	10	32	99.7376	0.886115
20	5	32	99.6185	0.800823
30	10	32	99.7402	0.884453
10	**10**	**64**	**99.7467**	**0.893935**
10	20	32	99.7168	0.909515

ADVANTAGES OF RANDOM FOREST

- It has an effective method for estimating missing data and maintains accuracy when a large proportion of the data are missing.
- Random Forest machine learning algorithms can be grown in parallel.
- Reduction in overfitting: by averaging several trees, there is a significantly lower risk of overfitting.
- Less variance: By using multiple trees, you reduce the chance of stumbling across a classifier that doesn't perform well because of the relationship between the train and test data.
- Has higher classification accuracy.

PERFORMANCE COMPARISON